Effective Metadata Systems for Archives

EAD and Unified Collection Management Systems

Gregory Wiedeman

University Archivist

University at Albany, SUNY

Archival Metadata Systems

- Paper Finding Aids
- HTML Finding Aids
- EAD records
- Accession Databases
- Unified Collection Management Systems
 - Archivist's Toolkit
 - Archon
 - ArchivesSpace
 - Proprietary or Museum-focused systems

Functions of Archival Metadata Systems

- Creation
- Management
 - Editing & Modifying
 - Enable Administrative functions
- Validation
 - Error-checking
- Public Access
- Aggregation
- Future Migration

Anne J. Gilliland-Swetland, "Popularizing the Finding Aid: Exploiting EAD to Enhance Online Discovery and Retrieval in Archival Information Systems by Diverse User Groups," *Journal of Internet Cataloging* vol. 4, no. 3/4 (2001)

Local Habits Die Hard

- Flexibility of finding aids
- Archives unique, demand unique practices?
- Local practices doesn't mean bad decision-making
 - Unique practices come from unique local context
- Problem was standardizing practices at the local level

Giving Data Structure

- Paper finding aids have no structure
 - Not machine readable
 - Not able to have automated functionality
- Structured data is unitized, has predefined rules
- Necessary to develop tools
- Use of XML to give structure to hierarchical archives metadata
 - EAD Standard

Problems with EAD, a Plural Standard

- Allowance of mixed content and Unstructured Data
 - <physdesc> 12 cubic ft.</phydesc>
 - <langmaterial>This collection is in <language>English</language>.</langmaterial></language>.</langmaterial></language>.</langmaterial></language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</language>.</l
- Failure to completely differentiate storage and display
 - Complains about public accessibility
- Difficulty of validation
 - Inconsistencies in valid EAD files
- Flexibility/Permissiveness
 - Many different ways to encode the same thing
 - Obstructs aggregation, migration
 - Makes overhauling the standard difficult
 - Hinders the development of modern online access tools

Problems with EAD, a Plural Standard

- Allowance of mixed content and Unstructured Data
 - Influenced by TEI
- Failure to completely differentiate storage and display
 - Web community still implementing concept
- Difficulty of validation
 - Problem with XML as a whole
- Flexibility/Permissiveness
 - Key to the wide use of EAD

Elizabeth H. Dow, "Encoded Archival Description as a Halfway Technology," *Journal Of Archival Organization*, 7 (2009)

Still advantages to EAD

- Can be locally standardized with work
- As close to digitally universal as you can get
- Hierarchical structure matches archives metadata
- New tools and cheap processing power to take advantage of XML
- Raised the technological skills of archivists

- Manage Administrative and Descriptive Metadata
 - Specific to archives
- Graphical User Interfaces
- Actually a way to avoid EAD
 - Rely on relational databases
 - Allows for more standardization
 - Can usually export to EAD

- Archivist's Toolkit (no longer supported)
 - Focus on back-end accessioning and creation of description
- Archon (no longer supported)
 - Focus on front-end online access as well as the creation of description
- AtoM (Access to Memory)
 - Front end online access system
- ArchivesSpace
 - Goal to be complete end to end archival metadata system
 - SS code4lib article issue 9, 2010-03-22

ArchivesSpace

- Archivist's Toolkit backend
- Archon public access
- Multiple systems redundant?
 - Combine resources
 - Share development effort
- Model is important, not software
 - One unified system

- Can often be stricter than EAD
- Can abrade with local practices
- May need to make minor adaptions to your local workflow
 - Many arbitrary practices
- Overall, wide use of these tools will standardize practices
 - Not always comfortable
- Often promotes better metadata creation

The Persistence of EAD

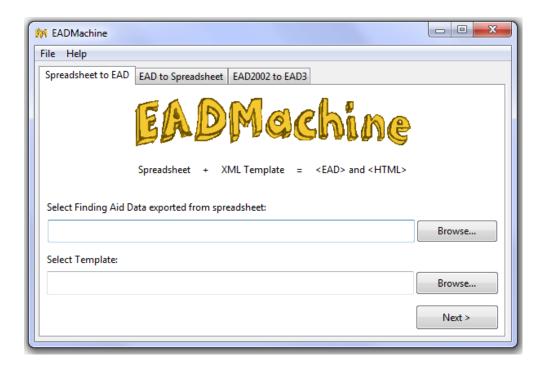
- Effective as Preservation Standard
- Yet EAD is still at the center of complex workflows
 - Automated, not manual
- EAD as a Metadata Dump
 - Unifying systems
 - Single Access point

Developing Better Descriptive Metadata at UAlbany

- No Unified Collection Management Tool
- Finishing full implementation of EAD
- Inconsistent metadata
- Manage metadata before migration

Using EADMachine to Create Consistent EAD

Cut Calibr	- 11 - A A =	😑 😹 🐨 🔤 Wrap	Text	-	Normal	Bad	Good	Neutral	Calculation	^
Format Painter	z u • 🖽 • 🔕 • 🗛 • 🚍 🕸	■ == 字 字 🔤 Merg	e & Center * \$ * % •	+₂g .og Condition	nal Format	Explanatory	Input	inked Cell	Note	
Clipboard G	Font	Alignment	G Number	Formattin	g * as lable *		Styles			
A7 - (=	fx Collection Name									
A	В	C D	E	F	G	н	1		J I	ĸ
			Other IDs			Other Representa				
EA DIN	lachine					any HTML or PDF ver				
	acialitie	Local Type	ID		Link	Name	Show?			
Collecti	on Information									
ollection Name				1						
		Maintena	nce Information		General Finding Aid	Notes				
ubtitle			AD 3 only]	<a>agencyname	>					
ponsor		Maintenance Agency		Name of agency	maintaining the					
JRL nclusive Date		Agency Code		finding aid. Requ	Ired In EAUS					
nclusive Date nclusive Normal		Maintenance Status Publication Status	new published		-					
ulk Date		Publication status	Providence	1212						
ulk Normal				Revis	ion Information	1				
rocessed by			EAD2002 will use <cha< td=""><td></td><td>revisiondesc>, EAD3 will</td><td>use <maintenancehis< td=""><td>tory></td><td></td><td></td><td></td></maintenancehis<></td></cha<>		revisiondesc>, EAD3 will	use <maintenancehis< td=""><td>tory></td><td></td><td></td><td></td></maintenancehis<>	tory>			
rocessing Date		Event Type	Date	DateNormal	AgentType	Agent	Description			
rocessing Date Normal										
ublisher										
ddress Line 1	Publisher Address									
ddress Line 2										
ddress Line 3					Finding Aid Langu	ages				
ublication Date		1	Languages	Language Code		Script Code	Note			
ublication Date Normal										
inding Aid Author										
inding Aid Creation Date										
reation Date Normal										
art of Series Jumber in Series		Abbreviation	Citation	Convention Lin	Standardized Conv	entions Link Title	Actuate	Show	v Note	
dition		Abbreviation	citation	convention Lin	vermed	Link Hue	Actuate	51101	Note	
	Edition Notes									
					Local Conventions					
		Abbreviation	Citation	Convention Lin	k Link Title	Actuate	Show	Note		
									-	
			Local C	ontrols						
		Term	LocalType	Date	DateNormal					
Finding Aid Info										
P P Finding Aid Info //	Collection Info / Collection Info2	/ Hist + Scope Notes / I	Containers 🦯 😏 🦯				[] ◀ [



http://github.com/gwiedeman/EADMachine

Python Scripts to Manage EAD

- Unique ID System
- Remove special characters
- Standardize encoding
- Remove mixed content when possible
- Automate changes for digital objects
- Rule based-validation
- Separate public access system

http://github.com/gwiedeman/EAD_Tools

*C:\Users\gw234478\Documents\EAD_tools\uid.py - Notepad++									
File Edit Search View Encoding Language Settings Macro Run Plugins Window ?									
C 台 🖯 🕼 C 🖧 X 🐚 C D C # 🍇 ≪ ≪ 🖪 🔂 🗉 ¶ 🌉 🐼	1								
i uid.py 🔀									
1 # -*- coding: utf-8 -*-									
2 from lxml import etree as ET									
3 import StringIO 4 import os									
5 import string									
6 import sys									
7 reload(sys) 8 sys.setdefaultencoding('utf8')									
9									
10 inputdir = 'C:\\Users\\gw234478\\Documents\\UID\\findingaids'									
11 outputdir = 'C:\\Users\\gw234478\\Documents\\UID\\output\\' 12									
creator_list = 'C:\\Users\\gw234478\\Documents\\UID\\creator.xml'									
14 15 rep ID = "nam "									
15 rep_10 = nam_									
17 for input_file in os.listdir(inputdir):									
18 if input_file.endswith(".xml"): 19 input_xml = inputdir + "\\" + input_file									
20 parser = ET.XMLParser(remove_blank_text=True)									
21 FA_input = ET.parse(input_xml, parser) 22 FA = FA_input.getroot()									
FA = FA_input.getroot() pi = FA.getprevious()									
24									
25									
28 - pass									
29									
31 \Box if cmpnt.tag == "c01":									
32 $c1 = c1 + 1$									
33 cmpnt.set('id', "nam_" + coll_ID + "-" + str(c1)) 34 □ if cmpnt.find('c02') is None:									
35 - pass									
36 ⊡ else: 37 c2 = 0									
$\begin{array}{cccc} & & & & & \\ 38 & & & & & \\ 39 & & & & & \\ & & & & & \\ & & & & & \\ & & & &$									
40 c2 = c2 + 1 41 □ if cmpnt2.find('c03') is None:									
42 - (mppt2 set('id', "nam," + coll ID + "-" + str(c1) + "," + str(c2))									
$\begin{array}{c} \text{an price set (a + man_{-} + con_{-}cb + man_{-} + con_{-}cb + man_{-} + con_{-}cb + man_{-}cb + man_{-}cb$									
44 c3 = 0									

Modular Metadata System Design

- Multiple systems
- Move to ArchivesSpace for accession system
- Would need effort to migrate anyway
- Maintain separate public access system
- Possible continue to use EADMachine for now

Consistency, Consistency, Consistency

- Inconsistent application of tools undermines there use
 - Doesn't matter how good the tool is
- Need to focus on use of tools, than functionality of tools
- Consistency of Metadata, not tools
 - Do we really need one, single unified tool?
 - Benefits of shared resources?

Effective Metadata Systems for Archives

EAD and Unified Collection Management Systems

Gregory Wiedeman

University Archivist

University at Albany, SUNY